AD_____

Award Number: W81XWH-11-1-0119

TITLE: RNA Chimeras as a Gene Signature of Breast Cancer

PRINCIPAL INVESTIGATOR: Dr. Dezhong Liao

CONTRACTING ORGANIZATION: Regents of the University of Minnesota
Minneapolis, MN 55455

REPORT DATE: June 2014

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
June 2014	Annual	15 May 13-14 May 14

4. TITLE AND SUBTITLE
RNA Chimeras as a Gene Signature of Breast Cancer

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-11-1-0119

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Dr. Dezhong Liao

Email: djliao@hi.umn.edu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Regents of the University of Minnesota
Minneapolis, MN 55455

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT
This project is set to test a hypothesis that breast cancer may express many RNA chimeras not only because there are fusion genes derived from chromosomal translocation or rearrangement but also because of abnormal trans-splicing of RNA transcripts. Many of these RNA chimeras may influence the behaviors of breast cancer via undiscovered mechanisms. One of the astonishing finding in 2012 fiscal year is the fusion RNAs formed between the nuclear RNA (nRNA) and mitochondrial RNA (mtRNA). Because such fusion, if it indeed exists in human cells, discloses a series of novel RNA processing mechanisms, in the past year we put much effort to analyze it and publish it. The major results from the work in the past year are summarized below.

15. SUBJECT TERMS- none listed

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			USAMRMC
U	U	U	UU	22	19b. TELEPHONE NUMBER *(include area code)*

Table of Contents

Introduction

This project is set to test a hypothesis that breast cancer may express many RNA chimeras not only because there are fusion genes derived from chromosomal translocation or rearrangement but also because of abnormal trans-splicing of RNA transcripts. Many of these RNA chimeras may influence the behaviors of breast cancer via undiscovered mechanisms. One of the astonishing finding in 2012 fiscal year is the fusion RNAs formed between the nuclear RNA (nRNA) and mitochondrial RNA (mtRNA). Because such fusion, if it indeed exists in human cells, discloses a series of novel RNA processing mechanisms, in the past year we put much effort to analyze it and publish it. The major results from the work in the past year are summarized below.

Body

As the only grant the PI (Dr. Liao) has since 2011, we have published or co-published with collaborators a total of 31 papers, seven of which are directly related to this project (1-7).

Trimeric RNAs, including those mtRNA-nRNA fusions, are finally published

Since last year's report, we have continued on identification of chimeras that contain mtRNA, i.e. those mtRNA-nRNA fusions, as outlined in the task 1 of the new Statement of Work (SOW) submitted last year. Because there are too many chimeric RNA sequence databases available for the public, we wrote a simple computer code to screen public databases, in a collaboration with Beijing Genomics Institute (www.genomics.cn), which is one of the world's largest genomic organizations. After tedious bioinformatic analysis, we published 61 trimeric expression sequence tags (ESTs), which are the cDNAs containing sequences from three different genes (the table below, copied from the appendix I), and 57 ESTs that contain mtRNA sequences (see table 3 in the appendix I). This publication (appendix I) proposes, for the first time, the possible existence of trimeric RNA and the fusion between mtRNA and nRNA. Working on additional data to address concerns of the reviewers and many other RNA experts took us unexpected efforts, which hampers our progress on other tasks.

Table 1. Trimeric ESTs.

AI335862	BF826714	BF826602	BF764896	BF762577	BF744644	BF331329	BF306729	BE694080
AI924910	AU132130	BF109407	AV744183	AV729389	AV725012	BE814336	BE762537	BE876577
AW608255	BE715872	BE715869	BE715858	BE709675	BE694009	BE696199	BQ348968	AA514694
AW956968	BF803049	BF764896	BF331329	AU142287	BE172179	X93499	BM824189	BG995785
AW999004	BQ689257	BQ689139	BU539467	AI925024	BF814512	BG003110	R19361	BE898652
BC064904	M77198	BM915020	BI004882	BF995070	BF878278	BF109407	AW994480	BE716966
BE074730	BE876742	BE937759	BF987118	BM691077	BM703781	BX109950		

Note: Each of these ESTs contains three sequence elements.
doi:10.1371/journal.pone.0077016.t001

Some fusion ESTs contain poly-A between the two partners, which together with mtRNA-nRNA suggests new RNA splicing mechanisms

Intriguingly, in some chimeric and trimeric ESTs, the gap sequence, which is an unmatchable sequence inserted between the two neighboring partners, was actually a poly-A sequence, or a poly-T when the upstream partner was reverse-complementary to the mRNA. This feature indicates that the downstream partner was fused to the poly-A tail of the upstream partner, such as in the EST AU142287 in which the mt-sequence was fused to the poly-A tail of the upstream nRNA (Fig 1, copied from appendix I). The poly-A or poly-T was usually appended to an earlier termination, but not the canonical end, of the last exon, such as in the ESTs AW608255 and AU142287. In some other cases such as the BM691077, the poly-A was appended at the 3' end of

the mtRNA sequence; more likely, the polyadenylation of the mtRNA occurred before fusion by the nRNA element.

If there are mtRNA-nRNA fusions, even if there is just a single one, it implies that either nRNAs are transported into the cytoplasm, likely into the mitochondria, to fuse with mtRNA, or mtRNAs are transported into the nucleus to fuse with nRNAs. Appearance of mtRNA in the nucleus has been reported (8), but it is technically difficult to confirm because the nuclear chromosomes contain as many as 755 mitochondrial pseudogenes (9, 10). Appearance of nRNA in the mitochondria has also been reported (11-14), although more studies are needed to confirm that large nRNAs are as easy as microRNA to relocate to the mitochondria. No matter which one goes into which, it indicates a previously unknown mechanism for RNA fusion, besides the transcriptional-slippage and trans-splicing, because transportation of nRNA to the mitochondria or fusion of the mtRNA to the nRNA should occur after the nuclear transcription has been terminated and splicing has been completed. The identification of those ESTs in which the downstream partner is fused to the poly-A tail of the upstream partner also supports a "post-transcription and post-splicing" mechanism, because polyadenylation usually occurs after both RNA transcript cleavage and splicing completion. Moreover, this new mechanism may occur outside the nucleus, whereas both transcriptional-slippage and trans-splicing of nRNA can only occur in the nucleus.

Fig. 1: Poly-A as the gap between two fusion partners. **Left panel**: In a fusion RNA, two neighboring genes' sequences have three relationships, i.e. 1) with an overlapped sequence, 2) with an unmatchable sequence as a gap, or 3) directly joined. **Right panel**: The 1-160th nt of EST AU142287 are matched the 429-588th nt of the last exon of the POLDIP3 mRNA from chromosome 22, followed by a poly-A tail (boldfaced and underlined). Since the wild type form of this exon should have 2248 nt, the polyadenylation actually follows an early transcriptional termination. The 178-526th nt region (italicized grey) after the poly-A is part of the ND2 mRNA from mtDNA. The 527-836th nt sequence belongs to the first two exons of the GNB2L1 mRNA from chromosome 5, with the first 10 nt (underlined lowercase letters) alternatively initiated from the -10 bp of the GNB2L1 gene. The last 85 nt (underlined) have a few deletion mismatches to the first 88 nt of exon 2 of the GNB2L1. The last 5 nt (agngg) are unmatchable and might belong to the cloning vector.

Table 5. Frequencies of three different relationships between two neighboring parnters in chimeric, trimeric and tetrameric RNAs.

Database	Relationship	N	Mean	Median(Q2)	Maxima(Q4)	Q1	Q3	Sstdv	Mad
mRNA in NCBI	Exact junction	157							
	Overlapped	736	6.5	3	606	2	6	29.52	2.97
	Gap (insert)	373	132.4	49	1303	13	167.5	198.79	65.23
EST in NCBI	Exact junction	2520							
	Overlapped	23252	6.4	5	190	3	7	7.7	2.97
	Gap (insert)	8998	29.5	21	307	8	38	31.4	20.76

Note: Putative chimeric, trimeric and tetrameric RNAs were identified from mRNA and EST collections of the NCBI database by our simple computer code. "Exact junction" means that the two partner mRNAs join directly. "Overlapped" means that the two partner mRNAs have at least 1 nucleotide (nt) overlapped. "Gap (insert)" means that there is at least 1-nt or an unmatchable sequence inserted between the two partner RNAs. The length of the overlapped or inserted sequences is short in most cases and is not normally distributed. Therefore, the mean, midian (Q2) and maximal lengthes in number of nt are calculated, besides the length at the 25%th (Q1) and 75% (Q3). Mad(X) = 1.4826 * median (|Xi-median(X)|). * When the distribution is so different from normal distribution, we usually compare "mean" to "median" and "standard deviation" to "mad". If the difference between mean and median, STDEV and MAD are huge, the distribution is far from normal distribution.
doi:10.1371/journal.pone.0077016.t005

Some possible reasons for artificial chimeric RNAs are proposed

It was noticed before that most chimeric RNAs have a short homologous sequence (SHS) shared by the two partner RNAs. This overlapped sequence was hypothesized to be a mechanism for the formation of chimeric RNAs at the RNA level either via transcription slippage or trans-splicing (15). Indeed, our data show that 736 (58.1%) of 1266 chimeric mRNAs or 23252 (66.9%) of 34770 of ESTs in the NCBI database contain such SHS, which on average has 6.4-6.5 nucleotides (nt) overlapped by the two partner genes (see the table above copied from appendix I). However, while we conducted the revised task 1 by using reverse transcription and polymerase chain reaction (RT-PCR) to amplify and clone thousands of published chimeric RNAs, we failed to confirm the existence of the vast majority of them in any breast cancer or normal cell lines or tissues. During conduction of these cloning procedures, we realized that the SHS may serve as a primer, with its antisense as the template, to create a wrong-template extension during RT or PCR, thus creating an artificial chimera of cDNA (16). We have also realized another weakness of RT that may lead to generation of artificial chimeras, which to our knowledge has not been described before: After an RT primed by the intended (usually 5'-tagged) primer is finished and the RNA template has degraded or has been digested by the RNase-H activity of the reverse transcriptase, the resulting cDNA may anneal to a new RNA fragment at the 3'-end via a SHS, as illustrated in figure 2. This SHS will serve as the primer to initiate a second RT reaction that elongates the cDNA, thus creating a chimera. This is possible since a retrovirus uses cellular tRNA to prime mRNA for reverse transcriptases to synthesize the first DNA strand (17, 18) and reverse transcriptases are known to be able to utilize endogenous small RNAs to efficiently prime cDNA synthesis in vitro (19-21). Since pentamers are often used in PCR, we assume that the random annealing only requires a SHS as short as five nucleotides, resulting in a chimeric cDNA in which the two partners share this SHS. This "consecutive RTs" scenario enlightens us in that many SHS-containing chimeric RNAs obtained by RT-based approaches may simply be technical artifacts (16).

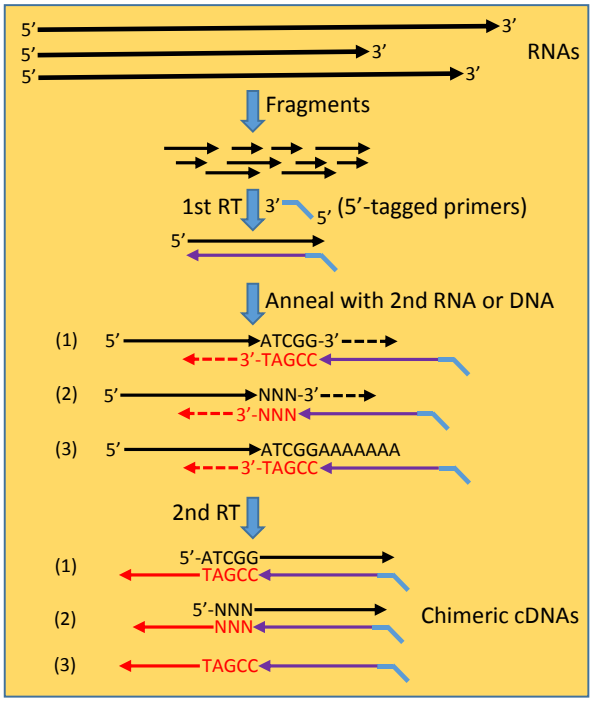


Fig. 2. The "consecutive RTs" hypothesis for formation of spurious chimeric RNAs. In a typical procedure of sample preparation for RNA-sequencing, RNAs are fragmented, usually by metal ion hydrolysis, to about twice the expected read size, resulting in a large number of RNA fragments. RT reaction, usually primed by 5'-tagged random primers, engenders the 1st cDNA strand. After the RNA template has degraded or has been digested by the RNase-H activity of the reverse transcriptase, the cDNA may anneal to another RNA fragment at the 3' end by five or more nucleotides, say ATCGG/TAGCC (scenario 1) that is referred to as "short homologous sequence" (SHS). This annealing initiates a second RT, producing a chimeric cDNA in which the two partners are joined at the SHS. In many cases, reverse transcriptase may append, at the cDNA end in a non-template manner, one or several nucleotides, usually CCC or GGG but also other base or bases, thus indicated as "NNN" (scenario 2). In this situation, the resulting chimera has the two partners joining with a shorter SHS or even without a SHS (when five or more bases are appended that alone constitute the primer). DNA residuals in the RNA samples, especially in the cases where the RNA samples have not been DNased, may also be molten to single-stranded oligos, which then anneal to the cDNAs via a SHS, resulting in a DNA-cDNA hybrid, because reverse transcriptase has DNA-

dependent DNA polymerase activity, i.e. can use DNA as the template. Annealing to the second RNA or DNA can occur at any place of the molecule before the poly-A or poly-dT tail (scenario3). If the second template is a single-stranded DNA oligo, the resulting chimera is partially double-stranded since the RNase-H activity of the reverse transcriptase cannot remove DNA template. This "consecutive RTs" scenario may also occur in routine RT reactions.

RNA samples usually contain DNA residuals, especially mitochondrial DNAs that are small and in a circular structure, because one cell has hundreds or even thousands of mitochondria. DNase treatment of the RNA sample can decrease the amount of, but usually cannot completely remove, DNA residuals (2, 6). Some of these DNA fragments may be molten to single-stranded oligos and serve as templates. The cDNA may also anneal to these single-stranded genomic or mitochondrial DNA oligos, resulting in DNA-cDNA chimeras in the second RT reaction, because reverse transcriptase from MMLV has DNA-dependent DNA polymerase activity, i.e. can use DNA as the template (22-24).

It is worth mentioning that the "consecutive RTs" scenario described above can also occur in routine RT, and many chimeric ESTs may be technical artifacts so derived. Moreover, if two RNAs may be artificially fused in this way, so can three or more RNAs as well; some of the trimeric or even tetrameric ESTs we identified may be such artifacts. However, the number of RNA fragments in routine RNA samples is smaller, and the resulting cDNA fragments are fewer; therefore there are fewer anneals to occur, relative to the RT during the sample preparation for RNA sequencing that involves RNA fragmentation. This inference, for the first time, reminds peers of a possible pitfall of today's RNA deep sequencing technology that is recently spread swiftly.

We also realized that RNA or DNA samples contained a lot of small RNA or DNA fragments, which should not be surprising because many RNA or DNA molecules must be degraded by RNases or DNases or during sample preparation. These RNA or DNA shards could serve as endogenous random primers (ERP) during RT or PCR. The existence of ERP and the SHS together set complex pitfalls and create a large number of artifacts, including spurious chimeras of RNA, during our routine molecular cloning, which unfortunately have not been well aware before by peers and us. Gene-specific primers have been widely used in cloning and expression studies for decades, but our results suggest that, because of the presence of ERPs, there basically is no such thing called "gene-specific primer" or "strand-specific primer" in RT-PCR based RNA cloning or sequencing.

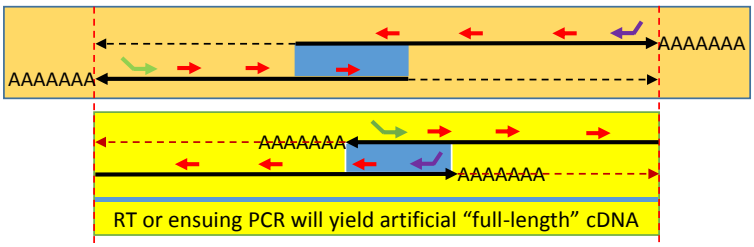


Fig. 3. Spuriousness caused by endogenous random primers (ERPs): When an antisense is also expressed and partly overlaps (blue area) with the sense transcript at either the 5'-end (top panel) or the 3'-end (bottom panel), ERPs (short red arrows) will prime both RNA transcripts in an RT. If the two oppositely-oriented RNAs overlap at the 3'-end, each RNA can serve as an endogenous primer to covert the opposite RNA strand to cDNA during RT, resulting in an artificial "full-length" cDNA of either gene, similar to the result from an RT using poly-dT primer. When the two RNAs overlap at the 5'-end, the same "full-length" cDNA will still be made if PCR ensues. RT with a gene-specific primer (green or purple arrow), usually 5'-tagged, can still specifically convert the targeted RNA transcript to cDNA. However, the oppositely-oriented transcript, along with numerous other RNA transcripts expressed in the cells, will be simultaneously primed by ERPs and thus converted to cDNAs.

mtRNA-nRNAs seem to occur at the RNA level without a genomic basis

We took a great effort on task 2 of the revised SOW to determine whether any mtRNA-nRNA is associated with a genomic RNA in the cell nucleus. We studied 10 breast cancer cell lines and

15 breast cancer samples but did not find any corresponding fusion DNA. Although it is difficult to prove a negative result, we tend to conclude that at least the majority of mtRNA-nRNAs may be formed at the RNA level, but not derived from a fusion genomic DNA in the nucleus. In conduct of task 3 of the revised SOW, we tried to determine whether there is any mtRNA-nRNA expressed in breast cancer cell lines and tumor samples, but so far we have not found any of such chimeric RNAs that shows a high expression in any of the cell lines or tumor samples.

One fusion gene produces multiple RNAs:

Work in the past year did confirm many chimeric RNAs derived from fusion genes. It is so common that the same two partner genes can form different fusion genes and that the same fusion gene can also produce different RNA variants. During our conduction of the project we see so often these features, which surprises us although they can find examples in the literature, such as the BCR-ABL fusion genes and their mRNA transcripts from the well-known Philadelphia chromosome. For instance, we have cloned three BCAS4-BCAS3 chimeric RNAs from MCF7 breast cancer cells and published two of them (7). Several published studies have also reported BCAS4-BCAS3 fusion RNAs, but none of the reported sequences matches one another or matches any of the three sequences we cloned. Thus, there are many BCAS4-BCAS3 variants cloned by us and by others, which exemplifies the diversity of chimeric RNAs formed between the same two partner genes.

Fusion genes or fusion RNAs in breast cancer are not recurrent:

We found that unlike the situation in lymphomas and prostate cancers, fusion genes and fusion RNAs rarely have reoccurrence in breast cancer. The majority of about four hundred fusion RNAs we verified from Skotheim's list (25) are found only in one cell line or one breast cancer tumor tissue. For instance, the above mentioned BCAS4-BCAS3 fusion is only found in the MCF7 cell line. On the other hand, all cell lines and tumor tissues we studied expressed at least three chimeric RNAs. Several published studies also report similar phenomena. One study reports that a RPS6KB1-VMP1 fusion occurs in about 30% breast cancers (26), but other similar studies do not find this fusion (27-30). This incongruity is likely because one single study has the same artifact from the same experimental conditions such as the same set of primers. These two features, i.e. all cases express several chimeric RNAs but none of the chimeras has recurrence, are a bad omen as it makes it difficult to use a common regimen for the cancer treatment. On the other hand, however, these features may also be used to direct personalized treatment of the cancer. Although there are some published studies showing that some chimeric RNAs are formed at the RNA level via such as trans-splicing, we were not able to confirm the existence of any of these RNAs in any breast cancer cell lines or tissues. We thus conclude that only very few chimeric RNAs formed at the RNA level are authentic, while the vast majority that are not associated with a fusion gene are artifacts derived from such as the "consecutive RTs" described above.

Proteomics data provide inklings of the existence of fusion proteins

At least some chimeric RNAs should be translated to fusion proteins that differ in the molecular weight from each of the two partner proteins, but so far this cannot be tested for most of the putative fusion proteins because of a lack of antibodies that recognize only the fusion protein without a cross reaction to one or both of the partner proteins. Due to this huge constraint, we first used polyacrylamide gel electrophoresis (PAGE) with sodium dodecyl sulfate (SDS) to fraction proteins from HEK293 human embryonic cells (as controls) and MBA-MB231 breast cancer cells, followed by liquid chromatography–mass spectrometry (LC-MS/MS) to identify the proteins in a narrow gel stripe excised at 40-kD and 26-kD, respectively. LC-MS/MS identified 57.75% and 21.13% of the peptide spectra from HEK293 and MB231 cells, respectively, with the remaining

42.25% and 78.87% unmatchable. Surprisingly, only 24.5% and 36.2% of the identified proteins from HEK293 and MB231 cells, respectively, migrated as expected from their formula molecular weight (FMW), when a 10% divergence, i.e 36-44 kD or 23-29 kD, is considered as the wild type (wt) range for the 40-kD and 26-kD stripes, respectively, as shown in the table below. About 8.4% (from HEK293) or 26.0% (from MB231), and 67.1% (from HEK293) or 37.8% (from MB231), of the proteins migrated at sites larger or smaller, respectively, than their FMW, thus referred to as "larger" or "smaller" groups. In HEK293 cells, 42.1% of the proteins with FMW >44-kD might have two smaller isoforms, as they appeared at both 40-kD and 26-kD. Many proteins had large regions lacking any LC-MS/MS identified sequence, indicating that these regions are not actually expressed in the proteins. Thus, of the proteins at these SDS-PAGE positions, only about 1/4-1/3 are the wt. The remaining hefty majority of the LC-MS/MS identified proteins are either isoforms of the known proteins or fusion proteins containing known peptide sequence. Moreover, some of the unannotated peptides may also be fusion proteins, although most of them are unmatchable due to other reasons such as mutations. Thus, the whole isoform spectrum, which may include fusion proteins, may be a good biomarker for many biological aspects such as for cancer and cancer therapies. Moreover, these revelations also suggest that while antibody producers should be cautious when selecting those antibodies that can detect only the anticipated band on immunoblots, peers should be cautious when cutting away the unanticipated bands from immunoblots during manuscript preparation.

Distribution of LC-MS/MS identified proteins in the HEK293 and MB231 cells at the 40-kD and 26-kD stripes.

Stripe	FMW (kD)	Group	HEK293		MB231	In both HEK293 and MB231*		
			In this stripe	Also in the other stripe[#]		Proteins	of HEK293	of MB231
40 kD	>44	Smaller	660 (68.2%)	278 (42.1%)	142 (37.8%)	120	18.20%	84.50%
	36-44	wt	232 (24.0%)	93 (40.1%)	136 (36.2%)	98	42.20%	72.10%
	<36	Larger	76 (7.8%)	55 (72.4%)	98 (26.0%)	38	50.00%	38.80%
	total		968 (100%)	426 (44.0%)[$]	376 (100.0%)	256		
26 kD	>29	Smaller	725 (66.2%)	411 (56.7%)				
	23-29	wt	273 (24.9%)	12 (4.4%)				
	<23	Larger	98 (8.9%)	7 (7.1%)				
	total		1096 (100.0%)	430 (39.2%)[$]				

Note: #: the proteins identified in both stripes. *: The proteins that are identified in MB231 cells but can also be found in the 40-kD stripe from the HEK293 cells; the percentage is the 120, 98 and 38 proteins divided by the number of proteins in the same group of the indicated cell line in the same group. [$]: The 44.0% and 39.2% come from 426/968 and 430/1096, respectively.

With the same strategy, we also fractioned proteins from MCF7 and MB231 cells on SDS-PAGE and excised gel stripes at higher sites, i.e. at 72-, 55- and 48-kD. Similar to the above data from MB231 cells, only about 12-21% of the peptide spectra could be matched to the reference leading to protein identification, while the remaining 79-88% of the peptide spectra are unmatchable, as shown in the table below. Of the LC-MS/MS identified proteins, only about 22-36%, i.e. roughly 1/5-1/3, were in the wt range, whereas the remaining hefty majority were either larger or smaller than the FMW (Fig. 4). Again, some of the unmatchable peptides and some of the isoforms may actually belong to fusion proteins, although we still lack approach to prove it.

Petide spectra that are matched or are unmatchable to the reference

Sample	MB231			MCF7		
	Peptide Spectrum Matches (PSMs)	Search Input	Match Rate	Peptide Spectrum Matches (PSMs)	Search Input	Match Rate
72 kDa	9416	51414	18.31%	7624	45504	16.75%
55 kDa	11345	53899	21.05%	5640	48052	11.74%
48 kDa	12063	57078	21.13%	7228	43462	16.63%

Key Research Accomplishments

1. We have finally published the mtRNA-nRNA data, which, for the first time, suggest that mtRNAs, especially those transcribed from the non-coding regions of the L-strand, may be combined with nuclear RNA to enlarge the RNA repertoire. We found that human mtRNAs also undergo cis- and trans-splicing. We have also identified a novel cis-splicing derived mtRNA. These findings contribute significantly to the RNA biology (Appendix I).

2. We find that fusion RNAs occurs in breast cancer in a random manner and are hardly recurrent, in line with some reports in the literature. Even the same fusion gene gives rise to different fusion RNAs.

3. The PI (Dr. Liao), whose salary and position is supported by this grant, has published 31 scientific publications under this support in the past three years, of which 13 are after the last reporting period.

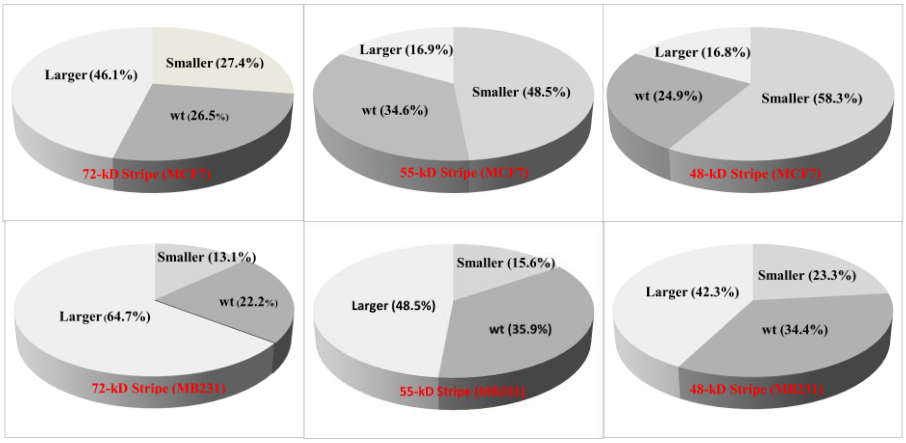


Fig. 4: LC-MS/MS identified proteins from narrow-stripes of SDS-PAGE gels excised at 72-kD, 55-kD and 48-kD, respectively. The wt proteins, defined as within 10% divergence from the site, i.e. 65-79 kD for the 72-kD band, 49-61 kD for the 55-kD band, and 43-53 kD for the 48-kD band, only constitute about 22%-36%, i.e. about 1/5-1/3, of the total proteins identified from MCF7 (top panel) and MB231 (bottom panel) cells. The remaining majority are either isoforms of known proteins or fusion proteins containing known peptide sequences.

Reportable outcome

1. A paper has been published on the trimeric RNAs and mtRNA-nRNAs. (Appendix I)
2. Many genes produce multiple protein isoforms, some of which may be translated to fusion proteins, as suggested by our LC-MS/MS data on narrow stripes from SDS-PAGE gels.
3. Chimeric RNAs in breast cancer are non-recurrent, basically.
4. One postdoctoral fellow is supported by the funds in the past year who coauthored the mtRNA-nRNA paper.

Conclusion

There probably have been over a million of putative chimeric RNAs reported in the literature or deposited in different databases. However, the vast majority of these chimeras remain unverified and therefore are still meaningless to us. Our results achieved so far lead to a conclusion that the majority of these putative chimeric RNAs may be technical artifacts. We have, in a publication, proposed major technical scenarios for how the artifacts are made, and a review article that better summarizes these scenarios has been submitted to a scientific journal upon the journal's invitation. We also conclude that most of those truly existing chimeras are associated with a fusion gene in the genome. Many genes produce multiple protein isoforms that constitute a surprisingly large portion of cellular proteins in both breast cancer cells and cells of non-breast origin, and some of the isoforms may be fusion proteins, although today we still lack a technology to prove it. However, in breast cancer, basically all chimeric RNAs are not recurrent, as having been reported

by others, and this feature emphasizes the importance of personalized diagnosis and treatment. Like nRNAs, mtRNAs also undergo cis- and trans-splicing and may fuse with nRNAs to enlarge the cellular RNA repertoire by forming some chimeric and trimeric RNAs, which implies a previously unaware mechanism for RNA fusion that may occur in the cytoplasm, but not in the nucleus.

References

(1) Liu B, Xu N, Man Y, Shen H, Avital I, Stojadinovic A, et al. Apoptosis in Living Animals Is Assisted by Scavenger Cells and Thus May Not Mainly Go through the Cytochrome C-Caspase Pathway. J Cancer 2013;4:716-23.

(2) Sun Y, Li Y, Luo D, Liao DJ. Pseudogenes as Weaknesses of ACTB (Actb) and GAPDH (Gapdh) Used as Reference Genes in Reverse Transcription and Polymerase Chain Reactions. PLoS One 2012;7:e41659-doi:10.1371/journal.pone.0041659.

(3) Sun Y, Sriramajayam K, Luo D, Liao DJ. A quick, cost-free method of purification of DNA fragments from agarose gel. J Cancer 2012;3:93-5.

(4) Sun Y, Luo D, Liao DJ. CyclinD1 protein plays different roles in modulating chemoresponses in MCF7 and MDA-MB231 cells. J Carcinog 2012;11:12-doi: 10.4103/1477-3163.100401.

(5) Sun Y, Lou X, Yang M, Yuan C, Ma L, Xie BK, et al. Cyclin-dependent kinase 4 may be expressed as multiple proteins and have functions that are independent of binding to CCND and RB and occur at the S and G 2/M phases of the cell cycle. Cell Cycle 2013;12:3512-25.

(6) Yang W, Wu JM, Bi AD, Ou-Yang YC, Shen HH, Chirn GW, et al. Possible Formation of Mitochondrial-RNA Containing Chimeric or Trimeric RNA Implies a Post-Transcriptional and Post-Splicing Mechanism for RNA Fusion. PLoS One 2013;8:e77016-doi: 10.1371/journal.pone.0077016.

(7) Yuan C, Liu Y, Yang M, Liao DJ. New methods as alternative or corrective measures for the pitfalls and artifacts of reverse transcription and polymerase chain reactions (RT-PCR) in cloning chimeric or antisense-accompanied RNA. RNA Biol 2013;10:958-68.

(8) Landerer E, Villegas J, Burzio VA, Oliveira L, Villota C, Lopez C, et al. Nuclear localization of the mitochondrial ncRNAs in normal and cancer cells. Cell Oncol (Dordr) 2011;34:297-305.

(9) Tsuji J, Frith MC, Tomii K, Horton P. Mammalian NUMT insertion is non-random. Nucleic Acids Res 2012;40:9073-88.

(10) Ramos A, Barbena E, Mateiu L, del Mar GM, Mairal Q, Lima M, et al. Nuclear insertions of mitochondrial origin: Database updating and usefulness in cancer studies. Mitochondrion 2011;11:946-53.

(11) Rorbach J, Minczuk M. The post-transcriptional life of mammalian mitochondrial RNA. Biochem J 2012;444:357-73.

(12) Mercer TR, Neph S, Dinger ME, Crawford J, Smith MA, Shearwood AM, et al. The human mitochondrial transcriptome. Cell 2011;146:645-58.

(13) Bandiera S, Ruberg S, Girard M, Cagnard N, Hanein S, Chretien D, et al. Nuclear outsourcing of RNA interference components to human mitochondria. PLoS One 2011;6:e20746-doi: 10.1371/journal.pone.0020746.

(14) Das S, Ferlito M, Kent OA, Fox-Talbot K, Wang R, Liu D, et al. Nuclear miRNA regulates the mitochondrial genome in the heart. Circ Res 2012;110:1596-603.

(15) Li X, Zhao L, Jiang H, Wang W. Short homologous sequences are strongly associated with the generation of chimeric RNAs in eukaryotes. J Mol Evol 2009;68:56-65.

(16) Yuan C, Liu Y, Yang M, Liao DJ. New methods as alternative or corrective measures for the pitfalls and artifacts of reverse transcription and polymerase chain reactions (RT-PCR) in cloning chimeric or antisense-accompanied RNA. RNA Biol 2013;10:958-67.
(17) Mak J, Kleiman L. Primer tRNAs for reverse transcription. J Virol 1997;71:8087-95.
(18) Marquet R, Isel C, Ehresmann C, Ehresmann B. tRNAs as primer of reverse transcriptases. Biochimie 1995;77:113-24.
(19) Colett MS, Larson R, Gold C, Strick D, Anderson DK, Purchio AF. Molecular cloning and nucleotide sequence of the pestivirus bovine viral diarrhea virus. Virology 1988;165:191-9.
(20) Gubler U. Second-strand cDNA synthesis: mRNA fragments as primers. Methods Enzymol 1987;152:330-5.
(21) Herzig E, Voronin N, Hizi A. The removal of RNA primers from DNA synthesized by the reverse transcriptase of the retrotransposon Tf1 is stimulated by Tf1 integrase. J Virol 2012;86:6222-30.
(22) Gubler U. Second-strand cDNA synthesis: classical method. Methods Enzymol 1987;152:325-9.
(23) Gubler U. Second-strand cDNA synthesis: mRNA fragments as primers. Methods Enzymol 1987;152:330-5.
(24) Spiegelman S, Burny A, Das MR, Keydar J, Schlom J, Travnicek M, et al. DNA-directed DNA polymerase activity in oncogenic RNA viruses. Nature 1970;227:1029-31.
(25) Lovf M, Thomassen GO, Bakken AC, Celestino R, Fioretos T, Lind GE, et al. Fusion gene microarray reveals cancer type-specificity among fusion genes. Genes Chromosomes Cancer 2011;50:348-57.
(26) Inaki K, Hillmer AM, Ukil L, Yao F, Woo XY, Vardy LA, et al. Transcriptional consequences of genomic structural aberrations in breast cancer. Genome Res 2011;21:676-87.
(27) Banerji S, Cibulskis K, Rangel-Escareno C, Brown KK, Carter SL, Frederick AM, et al. Sequence analysis of mutations and translocations across breast cancer subtypes. Nature 2012;486:405-9.
(28) Nik-Zainal S, Alexandrov LB, Wedge DC, Van LP, Greenman CD, Raine K, et al. Mutational processes molding the genomes of 21 breast cancers. Cell 2012;149:979-93.
(29) Shah SP, Roth A, Goya R, Oloumi A, Ha G, Zhao Y, et al. The clonal and mutational evolution spectrum of primary triple-negative breast cancers. Nature 2012;486:395-9.
(30) Stephens PJ, McBride DJ, Lin ML, Varela I, Pleasance ED, Simpson JT, et al. Complex landscapes of somatic rearrangement in human breast cancer genomes. Nature 2009;462:1005-10.

OPEN ACCESS Freely available online

PLOS ONE

Possible Formation of Mitochondrial-RNA Containing Chimeric or Trimeric RNA Implies a Post-Transcriptional and Post-Splicing Mechanism for RNA Fusion

Wei Yang[1]*, Jian-min Wu[1], An-ding Bi[2], Yong-chang Ou-yang[2], Hai-hong Shen[3], Gung-wei Chirn[4], Jian-hua Zhou[5], Emily Weiss[2], Emily Pauline Holman[2], D. Joshua Liao[2]*

1 Guangxi Veterinary Research Institute, Nanning, Guangxi, P.R. China, **2** Hormel Institute, University of Minnesota, Austin, Minnesota, United States of America, **3** School of life Sciences, Gwangju Institute of Science and Technology, Gwangju, Korea, **4** Biomedical Engineering Department, Boston University, Boston, Massachusetts, United States of America, **5** Nantong University, Nantong, JianSu, P.R. China

Abstract

Human cells are known to express many chimeric RNAs, i.e. RNAs containing two genes' sequences. Wondering whether there also is trimeric RNA, i.e. an RNA containing three genes' sequences, we wrote simple computer code to screen human expression sequence tags (ESTs) deposited in different public databases, and obtained hundreds of putative trimeric ESTs. We then used NCBI Blast and UCSC Blat browsers to further analyze their sequences, and identified 61 trimeric and two tetrameric ESTs (one EST containing four different sequences). We also identified 57 chimeric, trimeric or teterameric ESTs that contained both mitochondrial (mt) RNA and nuclear RNA (nRNA), i.e. were mtRNA-nRNA fusions. In some trimeric ESTs, the downstream partner was fused to the poly-A tail of the upstream partner, which, together with the mtRNA-nRNA fusions, suggests a possible new mechanism for RNA fusion that occurs after both transcription and splicing have been terminated, and possibly outside the nucleus, in contrast to the two current hypothetical mechanisms, trans-splicing and transcriptional-slippage, that occur in the nucleus. The mt-sequences in the mtRNA-nRNA fusions had pseudogenes in the nucleus but, surprisingly, localized mainly in chromosomes 1 and 5. In some mtRNA-nRNA fusions, as well as in some ESTs that were derived only from mtRNA, the mt-sequences might be cis- or trans-spliced. Actually, we cloned a new cis-spliced mtRNA, coined as 16SrRNA-s. Hence, mtDNA may not always be intron-less. Fusion of three or more RNAs to one, fusion of nRNA to mtRNA, and cis- or trans-splicing of mtRNA should all enlarge the cellular RNA repertoire, in turn enlarging the cellular functions. Therefore, future experimental verification of the existence of these novel classes of fusion RNAs and spliced mtRNAs in human cells should significantly advance our understanding of biology and medicine.

Citation: Yang W, Wu J-m, Bi A-d, Ou-yang Y-c, Shen H-h, et al. (2013) Possible Formation of Mitochondrial-RNA Containing Chimeric or Trimeric RNA Implies a Post-Transcriptional and Post-Splicing Mechanism for RNA Fusion. PLoS ONE 8(10): e77016. doi:10.1371/journal.pone.0077016

Editor: Arun Rishi, Wayne State University, United States of America

Received June 26, 2013; **Accepted** August 26, 2013; **Published** October 24, 2013

Funding: This work was supported by a grant from the Department of Defense of United States (DOD Award W81XWH-11-1-0119) to DJL. The funding agency had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exit.

* E-mail: gxyangwei@163.com (WY); djliao@hi.umn.edu (DJL)

Introduction

The swift spread of high throughput RNA sequencing technology in recent years has led to identification of tens of thousands of putative chimeric RNAs, which generally are considered as RNA molecules containing two genes' sequences. While a few hundreds of these chimeras are known to be transcribed from fusion genes that are formed due to genetic alterations, such as chromosomal rearrangements and genomic DNA deletion or amplification [1], for the vast remaining majority, how they are formed is still unknown. Trans-splicing may be one mechanism, but in mammalian cells it has not yet been well defined, in part because in mammals it is considered a rare event with a very different mechanism from that in low-level organisms [2]. Considering that the well characterized cis-splicing is a biochemical reaction that involves only one pre-mRNA as the substrate molecule and produces one mature mRNA, we define trans-splicing as a biochemical reaction that involves two RNA molecules as the substrates, no matter whether the product RNA codes for a protein or not. In other words, "one or two substrate RNA molecules" is a clear demarcation between cis- and trans-splicing, as illustrated in figure 1. The two substrate RNAs can be two copies of the same one; in this case trans-splicing often results in an RNA with duplicated exons, such as the 77–80 kD variant of human estrogen receptor alpha [3,4]. The two substrate RNAs can also be a sense and an antisense transcribed from the same genomic locus, and can be pre-mRNA of two different genes on the same chromosome or on different chromosomes. In a report from the ENCODE, it is estimated that RNA transcripts from about 65% of the human genes form chimeric RNA with a transcript from another gene, but in most cases this other gene is nearby on the same chromosome [5,6]. However, some of those chimeras formed between two tandem genes on the same chromosome may be derived from a single RNA transcript and thus are not real chimeras, by our definition. Instead, such a long transcript spanning two tandem genes can be considered as 1) an RNA product of a different, not-yet-annotated, gene, 2) an alternatively terminated transcript of the first gene, or 3) an

alternatively initiated transcript of the second gene. This long transcript may undergo a different cis-splicing, producing a mature RNA product, as illustrated in figure 1. Unfortunately, of those chimeras reported by the ENCODE or deposited in different public databases, it is unclear which ones are processed from a single RNA precursor and which other ones result from two separate precursors.

A significant percentage of putative chimeric RNAs encompass a short homologous sequence (SHS) shared by the two partners [7]. Although its reason is unknown, this feature has led to a so-called "transcriptional-slippage" hypothesis on how this type of chimeras are formed, which infers that two genes are in the same transcriptosome and transcription slips from one gene to the other, due to the complementarity at the SHS [7]. Since transcription of chromosomal DNA occurs in the nucleus and splicing occurs while the transcription is still elongating [8], both the transcriptional-slippage and the trans-splicing, if they indeed happen, should occur in the nucleus.

Since two RNA molecules could fuse to one, we wondered whether three RNAs could also fuse to one. In this study, we identified some expression sequence tags (ESTs) that contained three genes' sequences, thus coined as trimeric ESTs or trimeras. We also identified some ESTs that contained both mitochondrial (mt) RNA (mtRNA) and nuclear RNA (nRNA). In some ESTs, the downstream partner was fused to the poly-A tail of the upstream partner. Since mtRNA are not transcribed in the nucleus and polyadenylation occurs after both transcription and splicing are terminated, these findings are inklings of the existence of a previously unsuspected mechanism for RNA fusion, besides the hypothetic trans-splicing and transcriptional-slippage. The possible existence of trimeras and mtRNA-nRNA fusions in human cells suggests that the whole cellular RNA repertoire may be much larger than what we have known, which in turn enlarges cellular functions.

Results

There are trimeric and tetrameric ESTs

We designed simple computer code to screen preliminarily EST and mRNA collections in the NCBI database and obtained hundreds of putative trimeric ESTs. We then used the NCBI Blast and UCSC Blat browsers to analyze the sequences of these ESTs and identified 61 trimeric ESTs (table 1), such as AI924910 (Fig. 2). We also found that two ESTs, i.e. BE762537 and BQ638079 (Fig. 2), contained four sequence components, coined as tetrameric RNAs or tetrameras. There are many more ESTs that have passed our initial screening but have not yet been analyzed using Blast and Blat; hence, many more trimeric and tetrameric ESTs are anticipated.

Some trimeras or tetrameras, such as BQ638079, contained two or three nRNA elements that were derived from the same chromosomal locus but were not linear, i.e. not a cis-spliced product (Fig. 2); likely, trans-splicing was involved in formation of these fusion RNAs, if they were not artifacts. In many of these multi-component ESTs, at least one sequence element had multiple locations in the genome, either on the same chromosome or on different chromosomes (table 2), making it impossible to determine its true origin. For example, the first component, i.e. the most 5′ partner, of the AV725012 had tens of locations on quite a few chromosomes (data not shown), whereas in the BE074730 (Fig. 3) and BF109407, an nRNA element had several copies on the same chromosome. Some of the trimeras have putative open reading frame for protein translation, as analyzed using DNAstar software, but the open reading frame cannot be fully analyzed because most, if not all, ESTs have not yet been cloned for their full-length sequence.

nRNA may be fused to mtRNA

We also identified 57 ESTs that were nRNA-mtRNA fusion, i.e. contained both nRNA and mt-sequence (table 3). In most of these ESTs, the mt-sequence had one or more pesudogenes in the nucleus, which were more often referred to as NUMT (nuclear mitochondrial sequence) [9,10]. These 57 ESTs together contained a total of 74 NUMTs, of which 32 (43.2%) and 25 (33.8%) were localized in chromosomes 1 and 5, respectively, with the remaining 23% in other chromosomes.

Most of these mt-sequences were better matched to the mtDNA than to their nuclear counterparts, as exemplified by BF306729 (Fig. 4), and thus were more likely from the mitochondria than from the nucleus. At the positions where the mt-sequence was mismatched to the mtDNA or the corresponding NUMT, the mismatch was more often a deletion, as shown in BF306729 (Fig. 4). mtRNAs are known to be subjected to RNA editing, non-template 3′ addition, base modifications and other different posttranscriptional modifications [11,12], making its cDNA sequence often mismatched to the mtDNA and NUMT. Therefore, sometimes it is difficult to determine the origin of an mt-sequence in the ESTs by sequence alignment. On the other hand, there also were some ESTs in which the mt-sequence was identical to the corresponding NUMT, such as in the BE074730 (Fig 3); in this situation it is unsure that the mt-sequence is derived from the mtDNA, although currently there is little evidence

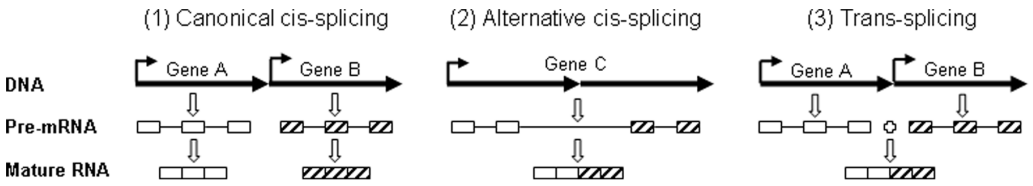


Figure 1. Our definition of cis- and trans-splicing in the situation of transcripts from two tandem genes on the same chromosome. (**1**): Two (A and B) tandem genes are transcribed separately as different pre-mRNA molecules, and each pre-mRNA is spliced independently to produce a mature RNA. This is canonical cis-splicing. (**2**): Transcription of gene A reads through the termination signal, due to whatever reason, and enters into gene B, resulting in a single, long pre-mRNA that can be regarded as: a) a transcript of a different gene (gene C), b) an alternatively terminated pre-mRNA of gene A, or c) an alternatively initiated transcript of gene B. This long pre-mRNA may be spliced to a different mRNA, such as one lacking the last exon of gene A and the first exon of gene B. We define this type of splicing of an alternatively transcribed pre-mRNA as an alternative cis-splicing. (**3**): Genes A and B are transcribed separately, and the two pre-mRNAs are spliced into one mature mRNA. We define this situation, if it happens, as trans-splicing. Hence, trans-splicing is a biochemical reaction involving two RNA transcripts as the substrate molecules, no matter whether they are transcribed from tandem genes on the same chromosome, from both Watson and Crick strands of the same genomic locus with a sense-and-antisense relationship, or from different chromosomes as two unrelated RNAs.
doi:10.1371/journal.pone.0077016.g001

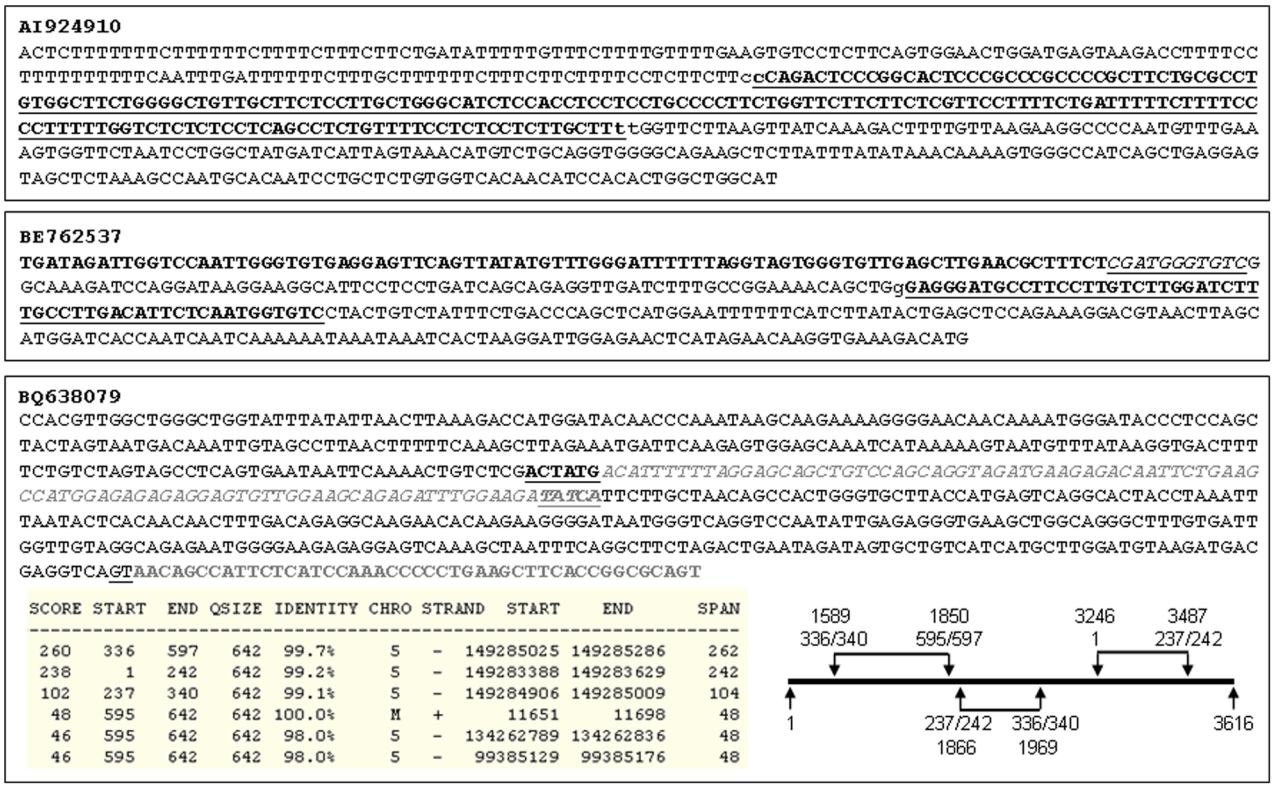


Figure 2. Examples of trimeric and tetrameric ESTs with or without an mt-sequence. Top panel: AI924910 is a trimera. Its 1–156th, 155–342th, and 341–550th nt regions match, respectively, the 5468–5676th nt region of the CYCO1 mRNA from chromosome 3, the 4332–4518th nt region of the PDCD11 mRNA from chromosome 10 (underlined and boldfaced), and the 366–521st nt region of the SREK1IP1 mRNA from chromosome 5. Between the two neighboring partners, there are two nt (lowercase letters) overlapped. **Middle panel**: BE762537 is a tetramera. Its first 86 nt (boldfaced) match the 2198–2283rd nt of the L-strand of mtDNA. The following 11-nt sequence (italicized and underlined) is an unmatchable gap. Its 98–168th nt sequence is part of the alternatively spliced exon 2 of the UBC mRNA from chromosome 12, which is followed by a 53-nt (the 268–220th) UBC antisense fragment (boldfaced and underlined). Both UBC sense and antisense fragments, which overlap at the 168th nt (the lowercase "g"), have multiple repeats in the UBC mRNA (not shown). The last 149-nt (the 221–369th nt) sequence is part of the ENPP6 mRNA from chromosome 4. **Bottom panel**: BQ638079 is a 642-bp tetrameric EST. Its first 595–597 nt sequence consists of three fragments from intron 7 of the PDE6A gene on the minus strand of chromosome 5. The relationships among these three fragments within this 3616-bp intron are illustrated in the figure, with the 3-digit numbers indicating the 5′ or 3′ end of each fragment and the 4-digit numbers indicating the position at the intron. As shown in the sequence, the figure, and a table copied from the UCSC browser, the middle fragment is 104-bp long, is matched to the 1866–1969th nt region of intron 7, and shares 6 or 5 nt (the underlined and boldfaced 237–242nd or 336–340th nt) with its up- or down-stream fragment that actually comes from the 3′ or the 5′ part, respectively, of the intron 7. The last 46–48 nt sequence of this EST come from the H-strand of mtDNA, with 2 nt (underlined GT in the sequence) overlapped with the last fragment of intron 7 of the PDE6A. The mt-sequence has two highly homologous NUMTs on chromosome 5, as shown in the table.
doi:10.1371/journal.pone.0077016.g002

suggesting that NUMTs are stably expressed. Several ESTs had a reoccurrence with identical sequence, which enhances their fidelity.

mtRNAs may undergo cis- and trans-splicing

Some chimeric and trimeric ESTs listed in table 3 that contained an mt-sequence showed one or several deleted regions that were set arbitrarily as 4 nucleotides (nt) or longer, such as in

Table 1. Trimeric ESTs.

AI335862	BF826714	BF826602	BF764896	BF762577	BF744644	BF331329	BF306729	BE694080
AI924910	AU132130	BF109407	AV744183	AV729389	AV725012	BE814336	BE762537	BE876577
AW608255	BE715872	BE715869	BE715858	BE709675	BE694009	BE696199	BQ348968	AA514694
AW956968	BF803049	BF764896	BF331329	AU142287	BE172179	X93499	BM824189	BG995785
AW999004	BQ689257	BQ689139	BU539467	AI925024	BF814512	BG003110	R19361	BE898652
BC064904	M77198	BM915020	BI004882	BF995070	BF878278	BF109407	AW994480	BE716966
BE074730	BE876742	BE937759	BF987118	BM691077	BM703781	BX109950		

Note: Each of these ESTs contains three sequence elements.
doi:10.1371/journal.pone.0077016.t001

BE074730

tgc*TAGCGATGATTATGGTAGCGGAGGTGAAATA*
*TGCTCGTGTGTCTACGT*GGTGGTGCTGGCGGCAG
GGAGCCAAAGTTCACCACAGGCAGCTGTGAGTCT
ACCATGGGTAGAAGCATCTGCTGGGCAGGAGCCC
CTGAAGGGAGAAAGCCACTTTGGCCACCAGGCTG
CAGAGGAGTAGAATAAAAATCCGAAGGGGATGGC
AGATCCTGGCGTACCTG**AAGCAACG**AAGTCTCAG
GTCTTTGTCCCTTGAAACCACTGGCTTTGGAGAA
ATCCTCAAACAGAAATGTGGAGAGTATCGGAATT
AAGGTTGTCATGGTGTGAACAGAGTAGATGATTG
CAGGAGTTCGAATCCAC

START	END	QSIZE	IDENTITY	CHRO	STRAND	START	END	SPAN	
52	221	357	100.0%	6_ssto_hap7	-	2935086	2935379	294	
52	221	357	100.0%	6_qbl_hap6	-	2897921	2898214	294	
52	221	357	100.0%	6_mann_hap4	-	2947168	2947461	294	
52	221	357	100.0%	6_dbb_hap3	-	2889856	2890149	294	
52	221	357	100.0%	6_cox_hap2	-	3113900	3114193	294	
52	221	357	100.0%	6	-	31604284	31604577	294	
52	221	357	99.5%	6_mcf_hap5	-	2983972	2984265	294	
230	357	357	100.0%	17	-	26653570	26653697	128	
231	356	357	92.9%	8	+	20791943	20792069	127	
4	51	357	100.0%	M	-	6797	6844	48	
4	51	357	100.0%	1	-	567347	567394	48	
4	51	357	98.0%	X	-	125605930	125605977	48	
4	51	357	98.0%	2	+	49456936	49456983	48	

Figure 3. BE074730 is a trimera. The origin of its 4–51st nt sequence (italicized and underlined) is unclear because it is completely matched not only to mtDNA but also to an NUMT on chromosome 1, as shown in the table copied from the UCSC browser. Its 52–221st nt sequence consists of two exons of the PRRC2A (also called BAT2) mRNA, with the downstream exon underlined; its real genomic origin is unclear, because this nRNA sequence belongs to a cluster of genes on chromosome 6 as shown in the table. The 230–357th nt sequence of this EST is part of the TMEM97 mRNA from chromosome 17. Actually, the 4–51st sequence also has homolog on chromosomes 2 and X, while the 230–357th sequence has homolog on chromosome 8. There is an 8-nt (boldfaced and italicized AAGCAACG) unmatchable gap between the 2nd and the 3rd partners.
doi:10.1371/journal.pone.0077016.g003

the AV702773 (Fig. 4), when aligned with mtDNA sequence. We also identified 12 ESTs, such as the AV729737 (Fig. 4), that contained only mt-sequence but the sequence had such deletions. Two of these 12 ESTs had reoccurrence (table 4). There were many other ESTs that were not counted because the deleted part was 3 nt or less.

In four mt-only ESTs, the mt-sequences were not linear along the mtDNA, thus likely being trans-spliced products. Besides the light (L) strand containing BE898652 shown in figure 4, the other three were the heavy (H) strand containing BF744644 and BF876577 and the BG995785 that contained both H- and L-strand sequences.

Some fusion ESTs contain poly-A between the two partners

In a fusion RNA, the two neighboring partners have three different relationships (Fig 5A), i.e. 1) with a sequence overlapped by the two partners, 2) with an unmatchable sequence as a gap, or 3) directly joined without a gap or an overlap. The frequency with an overlapped sequence was much higher than that with a gap, while the frequency for joining directly was the lowest (Table 5). In most ESTs in the databases we analyzed, the overlapped sequence was short, with 3–5 bp as the medium (table 5), just as coined by Li et al previously as SHS [7]. The gap sequence was also short in most cases, although much longer than the overlapped sequence (table 5). For unknown reasons, the gap was still unmatchable to any genomic region of human or any other organisms even when its length was long enough to be specific. Because it is unmatchable, it is regarded as a gap sequence, but not as an additional fusion element. Actually, we sometimes obtained such unmatchable sequence data during our cloning human cDNAs. If the hypothetical 3′-to-5′ inverse but not complementary sequences exist [13,14], it may be a possible explanation.

Intriguingly, in some ESTs the gap was actually a poly-A sequence, or a poly-T when the upstream partner was reverse-complementary to the mRNA, indicating that the downstream partner was fused to the poly-A tail of the upstream partner, such as in the AU142287 in which the mt-sequence was fused to the poly-A tail of the upstream nRNA (Fig 5). The poly-A or poly-T was usually appended to an earlier termination, but not the canonical end, of the last exon, such as in the AW608255 and AU142287. In some other cases such as the BM691077, the poly-A was appended at the 3′ end of the mtRNA sequence; more likely, the polyadenylation of the mtRNA occurred before fusion by the nRNA element.

Identification of 16SrRNA-s, a cis-spliced mtRNA

We performed reverse transcription (RT) and polymerase chain reaction (PCR) followed by T-A cloning and sequencing to verify those trimeric or terameric ESTs shown in figures 2, 3, 3, 5, i.e. the AI924910, BE762537, BQ638079, BE074730, BE898652, AV702773 and AU142287, using total RNA sample from the whole-cell lysate of HEK293 cells, but we failed to detect any of these seven ESTs. However, we serendipitously identified a cis-spliced mtRNA in HEK293 cells, which had a 769-bp, i.e. the 2067–2836th nt region of the mtDNA, deleted from the 16S

Table 2. ESTs that contain a sequence with mulitple repeats in nuclear DNA.

AK098472	BF826602	BF826714	BF803049	BF764896	BF762577	BF744644	EG328081	BF331329
BE172179	BF306729	AU132130	BF109407	AV751897	BF083272	BF083248	AV744183	AV727918
BE149219	AV725012	AV704911	AV703546	AV702773	BE899559	BE899119	BE898652	BE876577
DB370064	AV696226	AV695866	AV691857	BE814336	BE762537	BE716966	BE715872	BE709354
DB324119	BE709292	BE696199	BE694009	AV696226	BF803049	BF764896	BF378297	BE898652
DB277685	BF331329	BF185864	AU142287	BE898652				

Note: At least one of the sequense element in these ESTs has multiple copies in nuclear DNA.
doi:10.1371/journal.pone.0077016.t002

Table 3. Chimeric, trimeric or tetrameric ESTs that contain mitochondrial sequence.

	Mitochondria			Chromosome		Fusion
#	Access #	M strand/region	M-span (nt)	NUMT	Partner	
1	DB324119.1	L7212–7515	304	1	M-10	Chimera
2	BE898652.1	L8462–8527, L8569–8969	62, 401	1, 1	M-M-19	Trimera
3	AU142287.1	H4547–4895	349	1	22-M-5	Trimera
4	BF378297.1	L8443–8642	200	1	M-22	Chimera
5	BE716966.1	H7587–7790	204	1	M-1–11	Trimera
6	BE899119.1	H8399–8936	538	1	M-22	Chimera
7	AV744183.1	H4333–4399	67	1	M-19–9	Trimera
8	BF762577.1	L6897–7014, L9645–9779	118,135	1, 1	7-M-M	Trimera
9	BF764896.1	H11165–11248	84	5	2-M-6	Trimera
10	BE709292.1	L10705–10865	161	5	M-7	Chimera
11	BE709354.1	L10705–10864	160	5	M-7	Chimera
12	BE899559.1	H10677–1170	494	5	6-M	Chimera
13	BF083248.1	H14061–14231	171	5	10-M	Chimera
14	BF083272.1	H14061–14233	173	5	10-M	Chimera
15	AV751897.1	H10511–10874	364	5	M-19	Chimera
16	BF306729.1	H10637–10864	227	5	5-M	Chimera
17	BF744644.1	H15540–15721, H14092–14295	182, 204	5	M-M-11	Trimera
18	BE762537.1	L2198–2283	86	17	M-12–12–4	Tetramera
19	BE876577.1	H10689–11361, H84455–8680	493, 226	5, 1	M-M-11	Trimera
20	AV702773.1	H2639–3082	444	17	14-M	Chimera
21	AA514694	L9176–9208	33	1	M-8–8	Trimera
22	AA581515	L7399–7453	55	1, 17	M-17	Chimera
23	AA679609	L7399–7520	122	1	M-22	Chimera
24	AA679609	L7399–7520 (≈AA679609)	122	1	M-22	Chimera
25	AI925024	H2037–2239	203	3, 11, 5	14-M-13	Trimera
26	AW134795	L1619–1671	53	11	M-1	Chimera
27	AW370799	H8966–9070	105	1, 5	19-M	Chimera
28	AW753072	L14400–14455	56	18, 5, 17, 5	M-9	Chimera
29	AW821349	L12015–12081	67	5	M-16	Chimera
30	AW898803	H2647–2773	127	None	X-M	Chimera
31	AW950200	H7276–7520	245	1	5-M	Chimera
32	BE074730	L6707–6844	48	1, x, 2	M-6–17	Trimera
33	BE162186	H5117–5322	205	1	M-13	Chimera
34	BE876742	H7398–7457	60	1	19-M-1	Trimera
35	BE937759	L9080–9144	65	1	9-M	Chimera
36	BF852160	L4703–4822	120	1	17-M	Chimera
37	BF987118	H6568–6604	37	1	M-21–7	Trimera
38	BF988359	L6876–6938, L6950–7025	63, 76	1, 1	7-M-M	Chimera
39	BG995785	H10473–10589, L15168–15217	117, 50	5, 5	M-M-1	Trimera
40	BM691077	H14103–14159	57	5	17-M-1	Trimera
41	BM703781	H9567–9605	39	1	4-M	Chimera
42	BM997144	L7189–7520	332	1	M-2	Chimera
43	BP348380	H2852–3018	167	5	19-M	Chimera
44	BQ300150	H8936–8995	60	1	3-M	Chimera
45	BQ348968	L12338–12505	168	None	M-8–2	Trimera
46	BQ638079	H11651–11698	48	5, 5	5–5–5-M	Tetramera
47	CV385666	L2620–2837	218	11	M-7	Chimera
48	DA086571	H7201–7521	321	1	M-1	Chimera

Table 3. Cont.

	Mitochondria			Chromosome		Fusion
#	Access #	M strand/region	M-span (nt)	NUMT	Partner	
49	DA182598	H2417–2733	317	11,3,6,17,5	M-11	Chimera
50	DA365070	H1613–1672	60	11, 7, 5	M-7	Chimera
51	DA511096	H7192–7533	342	1	M-1	Chimera
52	DA757571	H2421–2762	342	None	M-1	Chimera
53	DB314922	L7393–7515	123	1	M-1	Chimera
54	DB324119	L7212–7515	304	1	M-10	Chimera
55	AW994480	L2654–2804, L2814–2880, L2894–2984	63, 67, 91	None	M-2–8	Trimera
56	BE694080	L2654–2984 (≈AW994480)	63, 67, 91	None	M-2–8	Trimera
57	BF826602	L11053–11116	64	5, 5	12–15-M	Trimera

Note: The first and last nt positions at the mtDNA of an mt sequence (M) are indicated, based on UCSC browser, while its length (span in the number of nt) may not always be calculated due to possible deletion of several nt. The chromosome or chromosomes that harbor an NUMT homologous to the mt sequence are indicated. The order of each partner in the chimera, trimera or tetramera is shown in the 5′-to-3′ orientation.

doi:10.1371/journal.pone.0077016.t003

rRNA, thus coined as 16SrRNA-s (Fig 6). The same mtRNA was also expressed in Hela cells, as confirmed by RT-PCR followed by T-A cloning and sequencing. PCR of total RNA sample from Hela cells without RT amplified the mtDNA, but not this 16SrRNA-s, indicating that it is not associated with an mtDNA with deletion (Fig. 6). Because it was detected by RT-PCR that lacks the strand-specificity, as described by us recently [15], it is unclear whether this novel mtRNA is transcribed from the H- or the L-strand of the mtDNA.

Discusion

Human cells may have trimeric or tetrameric RNAs

This study reports, for the first time, the existence of trimeric and tetrameric ESTs. Although our RT-PCR studies failed to confirm the expression of seven of these fusion RNAs in HEK293 cells, it cannot be excluded that these seven and other trimeras and tetramers are expressed in other situation or in other cell types. How the three or four different sequences are fused into one is unknown. It remains possible that the mechanism for formation of chimeras, trimeras and tetrameras is similar, no matter whether the fusion occurs as a cellular event or as a technical artifact.

Is mtRNA-nRNA fusion a mechanism to enlarge the RNA repertoire?

This study is also the first report on ESTs that are mtRNA-nRNA fusions. If this type of fusion truly exists and if the involved mt-element is truly derived from mitochondria, but not from its nuclear counterpart (i.e. NUMT), it suggests that two different pools of RNA, i.e. nRNA and mtRNA, can join together to constitute a previously unaware mechanism to enlarge the cellular RNA repertoire. Since most part of the transcript from the L-strand is known to be non-coding, it is possible that some parts of the L-strand transcript may be used as a resource for mtRNA-nRNA fusion to enlarge the ensemble of the cellular RNA, which is supported by a recent report of the existence of mtRNA antisenses [16]. Similarly, we found that the nRNA components in many ESTs were large and did not belong to exons of known genes. Likely, contribution to mtRNA-nRNA fusions may be a previously unsuspected role of non-coding nRNAs as well.

Considering that mtDNA encompasses only about 16.5 kb whereas the human genome contains 3164.7 million base-pairs (bp) (www.ornl.gov/sci/techresources/Human_Genome/project/info.shtml), it seems that mtRNA has a much higher frequency than nRNA to form fusion RNA, no matter whether the fusions are genuine or spurious. This may partly be because a set of processed mtRNAs are not polyadenylated [17], leaving them a higher chance to form spurious chimeras as we inferred recently [15].

mtRNA-nRNA suggests a new mechanism for RNA fusion

Tens of thousands of chimeric RNA have been identified so far, but the vast majority of them still remain putative, waiting for verification of their true existence and determination of how they are formed. For those true chimeras that are formed at the RNA level, i.e. are not transcribed from a fusion gene, transcriptional-slippage and trans-splicing are the two major hypothetical mechanisms. However, if there are mtRNA-nRNA fusions, even if there is just a single one, it implies that either nRNAs are transported into the cytoplasm, likely into the mitochondria, to fuse with mtRNA, or mtRNAs are transported into the nucleus to fuse with nRNAs. Appearance of mtRNA in the nucleus has been reported [18], but it is technically difficult to confirm because the nuclear genome contains as many as 755 NUMTs [9,10]. Appearance of nRNA in the mitochondria has also been reported [12,16,19,20], although more studies are needed to confirm that large nRNAs are as easy as microRNA to relocate to the mitochondria. No matter which one goes into which, it indicates a previously unknown mechanism for RNA fusion, besides the transcriptional-slippage and trans-splicing, because transportation of nRNA to the mitochondria or fusion of the mtRNA to the nRNA should occur after the nuclear transcription has been terminated and splicing has been completed. The identification of those ESTs in which the downstream partner is fused to the poly-A tail of the upstream partner also supports a "post-transcription and post-splicing" mechanism, because polyadenylation occurs after both RNA transcript cleavage and splicing completion, although the polyadenylation may continue in the cytoplasm [21]. Moreover, this new mechanism may occur outside the nucleus, whereas both transcriptional-slippage and trans-splicing of nRNA can only occur in the nucleus.

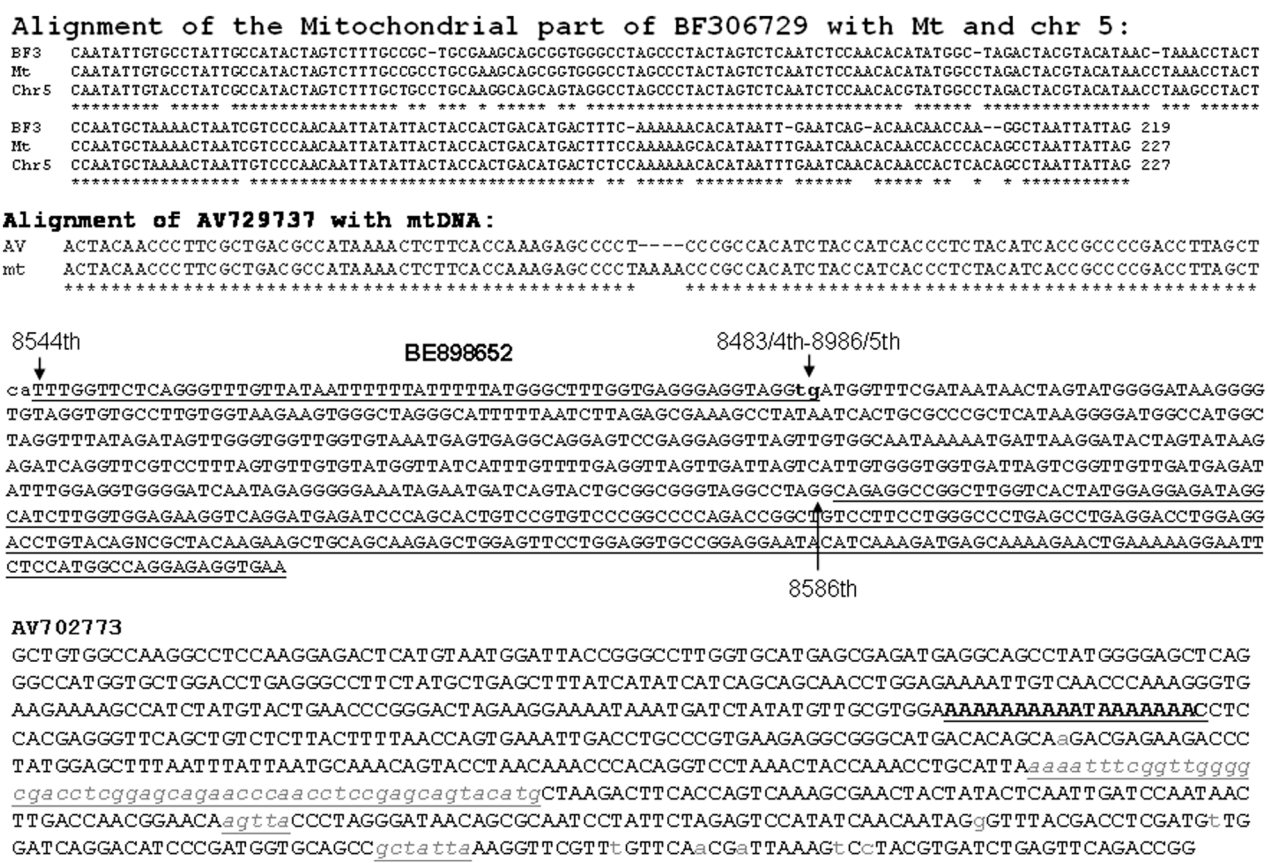


Figure 4. Cis- or trans-spliced mtRNA in ESTs. Alignment of the mt-part of EST BF306729 (BF3) with mtDNA (mt) and a corresponding NUMT on chromosome 5 (Chr 5) shows that the EST matches better to the mtDNA than to the NUMT, suggesting that the mt-sequence is more likely to be transcribed from mtDNA, but not its nuclear counterpart. Mismatched positions are often deletions. AV729737 is an mt-only EST that has an ''AAAA'' deletion when aligned with mtDNA. BE898652 is a trimera containing trans-spliced mt-sequence. Its 3–64th nt region is matched to the L-strand of an mtDNA (AC_000021), which is part of the ATP8 gene's antisense, while its 63–461st nt sequence is matched to the 8586–8986th nt of the L-strand, which is part of the ATP6 gene's antisense. The two sequences overlap at the boldfaced lowercase ''tg''. The last part of this EST (underlined) is part of the PSMC4 mRNA from chromosome 19. AV702773 is a chimera. Its 1–245th nt region belongs to the last 3 exons of the PSME2 mRNA from chromosome 14, which is followed first by a poly-A signal (the 246–263rd nt, containing a T) and then by an mt-sequence (265–708th nt) from the 16S rRNA gene (the 2639–3082nd nt of mtDNA). However, since the PSME2 mRNA (NM_002818.2) in the NCBI database does not have a poly-A signal or a poly-A tail, it is unclear whether this poly-A is part of PSME2 that is undocumented or is part of the following mtRNA that is polyadenylated. Note that one long and two short italicized lowercase sequences in grey within the mtDNA region do not actually exist in, i.e. are deleted from, AV702773. Several single lowercase letters in grey are mismatches to the mtDNA (AY195786.2).
doi:10.1371/journal.pone.0077016.g004

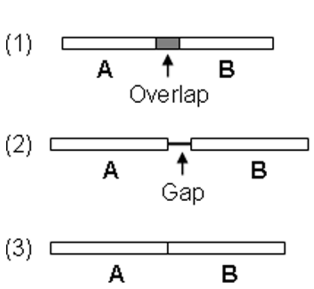
Do mtRNAs also undergo cis- and trans-splicing?

RNA editing as well as cis- and trans-splicing are known to occur in mtRNAs of plants and some low-level organisms [22–25]. However, to our knowledge there has only been one reported chimeric mtRNA in the human [26,27] and another one in the mouse [28]. Human mtDNA does not contain introns and thus

Table 4. ESTs that are mt-only sequences and are probably cis-spliced products.

EST	linear mt regions	EST	Linear mt regions
BE701368	H887–948, H1049–1108	BE932324	L651–1152, L1162–1218, L1223–1246
BE162186	H5117–5153, H5158–5322	BE932352	Reoccurence of BE932324
BG429734	multiple 4–14 nt gaps	BE875084	H5616–5792, H5800–5966, H5915–6171
BP348380	H2852–2921, H2829–3018	BE876577	H10869–11100, H11107–11361
AA070765	H1755–1881, H1888–2157	AV729737	H3017–3489, H3494–3721
AW994480	L2654–2804, L2814–2880, L2894–2984	BF676688	H4333–4912, H4933–5037
BE694080	Reoccurence of AW994480	BF988359	L6876–6938, L6950–7025

Note: These ESTs consist of only mt sequences but with one or more deletions of 4-nt or more, thus being cis-spliced products.
doi:10.1371/journal.pone.0077016.t004



AU142287
CAGTGTGTGCCTCCCCGAGCCTCAGCCCCAAGCTGATTTCTTATCTGGAAATGGTACACTGAATTCTCTGGGT
GGCTTTCTTGTGGCCCCATGGGATGCAGCGTGGGGGCTGTCTGAAGGACCCTGCTTTTTCCAGGGGCCGAGGG
GCTGCCTTTCCTTT**AAAAAAAAAAAAAAAA***GATTTTTTACCTGAGTAGGCCTAGAAATAAACATGCTAGCTT*
TTATTCCAGTTCTAACCAAAAAAATAAACCCTCGTTCCACAGAAGCTGCCATCAAGTATTTCCTCACGCAAGC
AACCGCATCCATAATCCTTCTAATAGCTATCCTCTTCAACAATATACTCTCCGGACAATGAACCATAACCAAT
ACTACCAATCAATACTCATCATTAATAATCATAATGGCTATAGCAATAAAACTAGGAATAGCCCCCTTTCACT
TCTGAGTCCCAGAGGTTACCCAAGGCACCCCTCTGACATCCGGCCTGCTTCTTCTCACATGACAAAAACTAGC
*CCCCATCTCAATCAT*gcttttctct*CTCTCTTTCACTGCAAGGCGGCGGCAGGAGAGGTTGTGGTGCTAGTTT*
CTCTAAGCCATCCAGTGCCATCCTCGTCGCTGCAGCGACACACGCTCTCGCCGCCGCCATGACTGAGCAGATG
ACCCTTCGTGGCACCCTCAAGGGCCACAACGGCTGGGTAACCCAGATCGCTACTACCCCGCAGTTCCCGGACA
TGATCCTCTCCGNCTCTCGAGATAAGACCATCATCATGTGGAAACTGACCAGGGATGAGACCAACTATGGAAT
TCCCAGCGGCTCTGCGGGGCACTNCCACTTTGTagngg

Figure 5. Poly-A as the gap between two fusion partners. Left panel: In a fusion RNA, two neighboring genes' sequences have three relationships, i.e. 1) with an overlapped sequence, 2) with an unmatchable sequence as a gap, or 3) directly joined. **Right panel**: The 1–160th nt of EST AU142287 are matched the 429–588th nt of the last exon of the POLDIP3 mRNA from chromosome 22, followed by a poly-A tail (boldfaced and underlined). Since the wild type form of this exon should have 2248 nt, the polyadenylation actually follows an early transcriptional termination. The 178–526th nt region (italicized grey) after the poly-A is part of the ND2 mRNA from mtDNA. The 527–836th nt sequence belongs to the first two exons of the GNB2L1 mRNA from chromosome 5, with the first 10 nt (underlined lowercase letters) alternatively initiated from the −10 bp of the GNB2L1 gene. The last 85 nt (underlined) have a few deletion mismatches to the first 88 nt of exon 2 of the GNB2L1. The last 5 nt (agngg) are unmatchable and might belong to the cloning vector.
doi:10.1371/journal.pone.0077016.g005

human mtRNAs are not supposed to undergo cis-splicing [17]. Therefore, our identification of the ESTs that contain two or more linear but segregated mt-sequences is a surprise, as it suggests that human mtRNA may also undergo cis-splicing. Such events may occur more frequently in disease situations since most ESTs are cloned from cancer cell lines, and since over 400 mutations or mtDNA rearrangements or deletions have been identified in different human diseases [29–31]. In the human mtRNA EU863789 identified by Burzio et al [26], the 2810–3125th nt region of the H-strand is trans-spliced to the 1673–3227th nt region of the L-strand, creating a 1866-nt chimeric RNA with its 311–316th nt region as a SHS shared by the two elements, according to our analysis. The EST BG995785 also contains both H- and L-strand sequences and thus may be a trans-spliced product as well. Another trans-spliced EST we identified is the EB898652, which is formed by two L-strand sequences (Fig. 4). Although most ESTs may be resulted from RT-PCR related techniques and thus are poor in the strand-specificity as we described recently [15], the interrelationship of the two mt-sequences within an EST should not be changed. The most convincing evidence for cis-splicing of mtRNA is our identification of the 16SrRNA-s in both HEK293 and Hela cells, as it is not associated with a deletion in the mtDNA (Fig. 6).

The mtRNAs that fuse with nRNA prefer to save copies in chromosomes 1 and 5

It is a surprising finding that the mt-sequences in mtRNA-nRNA fusions have their corresponding NUMTs mainly in chromosomes 1 and 5 (Table 3), because these two chromosomes do not have significantly more NUMTs than other chromosomes [9,32]. Since evolutional insertion of mtDNA into the nuclear genome is not a random event [10], a common mechanism, which has a preference to chromosomes 1 and 5, may regulate both the fusion between some mtRNAs and some nRNAs and the insertion of the same mtRNAs into the nuclear genome. This hypothetical thinking deserves further exploration.

Abundant mtDNA is an unvanquished obstacle for mtRNA detection

Because it is unclear whether the mtRNA-nRNA fusion occurs in the nucleus, cytoplasm or mitochondria, we determined the expression of seven fusion ESTs in the whole cell lysates with RT-PCR, but failed to confirm the existence of any of the seven in

Table 5. Frequencies of three different relationships between two neighboring parnters in chimeric, trimeric and tetrameric RNAs.

Database	Relationship	N	Mean	Median(Q2)	Maxima(Q4)	Q1	Q3	Sstdv	Mad
mRNA in NCBI	Exact junction	157							
	Overlapped	736	6.5	3	606	2	6	29.52	2.97
	Gap (insert)	373	132.4	49	1303	13	167.5	198.79	65.23
EST in NCBI	Exact junction	2520							
	Overlapped	23252	6.4	5	190	3	7	7.7	2.97
	Gap (insert)	8998	29.5	21	307	8	38	31.4	20.76

Note: Putative chimeric, trimeric and tetrameric RNAs were identified from mRNA and EST collections of the NCBI database by our simple computer code. ''Exact junction'' means that the two partner mRNAs join directly. ''Overlapped'' means that the two partner mRNAs have at least 1 nucleotide (nt) overlapped. ''Gap (insert)'' means that there is at least 1-nt or an unmatchable sequence inserted between the two partner RNAs. The length of the overlapped or inserted sequences is short in most cases and is not normally distributed. Therefore, the mean, midian (Q2) and maximal lengthes in number of nt are calculated, besides the length at the 25%th (Q1) and 75% (Q3). Mad(X) = 1.4826 * median (|Xi-median(X)|). * When the distribution is so different from normal distribution, we usually compare ''mean'' to ''median'' and ''standard deviation'' to ''mad''. If the difference between mean and median, STDEV and MAD are huge, the distribution is far from normal distribution.
doi:10.1371/journal.pone.0077016.t005

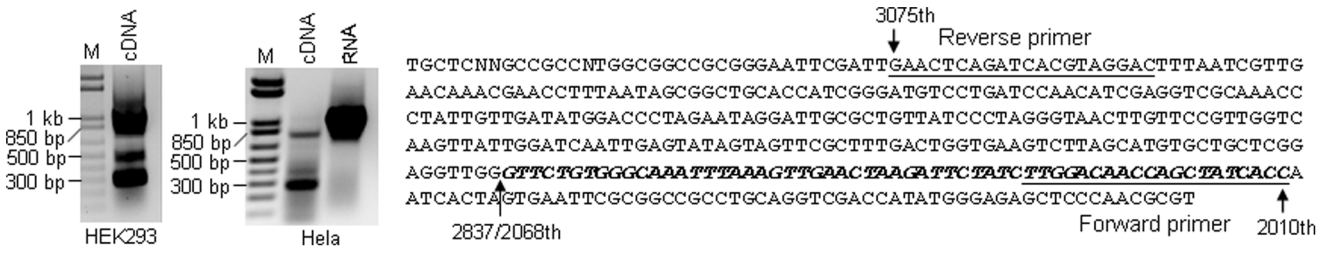


Figure 6. Expression of 16SrRNA-s. RT-PCR with primers that amplify the 2010-3075th region of mtDNA detects a 300-bp band and a 500-bp band, besides the anticipated band at about 1-kb, in HEK293 cells. The 500-bp band occurs randomly with greatly variable density, thus likely being a heterodimer of the 300-bp and 1-kb cDNAs, which appears randomly and is a common phenomenon as we have frequently described [37,39,40]. Isolation of the 300-bp band followed by T-A cloning and sequencing reveals that it is an mtRNA with a 768-bp (the 2069–2836th nt region of mtDNA) deletion, as indicated in the sequence. The boldfaced sequence is the downstream exon. The number of the nt is based on the UCSC browser, with the position of the last nt in the reverse primer (underlined) and the first nt in the forward primer (underlined) indicated. The sequences outside the primers belong to the T-A vector. Because the RNA sample was not pre-digested with DNase, the 1-kb band should be derived from not only cDNA, but also mtDNA, of the 16S rRNA gene. Removal of DNA from an RNA sample from Hela cells followed by RT-PCR (cDNA) still detects the 300-bp band, besides a band at about 850-bp, while the 1-kb band was very weak. Cloning and sequencing the 850-bp band reveal that it is part of the 16S rRNA that was amplified because the reverse primer is also partly reverse-complementary to the mtDNA. Direct PCR amplification of the RNA sample without RT (RNA) detects only the 1-kb band, but not the 300-bp one, indicating that the 300-bp band is not associated with an mtDNA with deletion.
doi:10.1371/journal.pone.0077016.g006

HEK293 cells. Study of mtRNA is technically more difficult than that of nRNA, partly because mtDNA is much smaller than chromosomal DNA and thus easier to mingle with cellular RNA. Moreover, normally one cell only has two sets of genomic DNA, although cancer cells are often of aneuploidy. In sharp contrast, one cell has from hundreds to thousands of mitochondria, and each mitochondrion contains multiple copies of mtDNA [33–35]. More complexly, many, but usually not all, copies of mtDNA may have mutations, deletions and other types of changes in cancer cells [29–31], which will likely cause additional bands in PCR amplification of mtDNA or mt-cDNA. Clearance of such huge amount of mtDNA by DNase followed by inactivation of the enzyme without causing RNA degradation is difficult, if not impossible. If DNase digestion is skipped, the mtDNA in the RNA sample will, to a large extent, compete out the cDNA of interest by depriving it from primers during PCR amplification of the RT products, which actually is somewhat reflected in figure 6. Moreover, because most mtRNA species are not supposed to undergo splicing and are thus identical in the length with their parental mt-genes, removal of mtDNA from the RNA sample is necessary in many situations, as we have discussed in more detail recently [36]. These technical constraints affect the RT-PCR detection of those mtRNA and mtRNA-nRNA fusions that are expressed at low abundance. We are trying to solve this technical hurdle so that we can better verify the expression of trimeric ESTs.

Conclusions

Sequence analyses of ESTs deposited in the NCBI suggest the possible existence of trimeric and tetrameric RNAs and the existence of mtRNA-nRNA fusions. The involved mt-sequences had their NUMTs preferentially in chromosomes 1 and 5. In some chimeric or trimeric ESTs, the downstream partner is fused to the poly-A of the upstream partner, which, together with the mtRNA-nRNA fusions, suggests a post-transcriptional and post-splicing mechanism for RNA fusion that does not necessarily occur in the nucleus, unlike the hypothetical transcriptional-slippage and trans-splicing. Moreover, we also identified many ESTs in which the mt-sequence might be a cis- or trans-spliced product, and we cloned a novel cis-spliced mtRNA coined as 16SrRNA-s, which together suggests that mtDNA may not always be intron-less. Fusion of several RNAs into one, fusion of nRNA to mtRNA, as well as cis- or trans-splicing of mtRNA all should enlarge the whole cellular RNA repertoire, in turn diversifying the cellular functions.

Materials and Methods

Computational sequence analyses

We wrote simple computer code and used it to screen preliminarily the EST and mRNA collections in the NCBI database to identify putative trimeric RNAs. We then used NCBI Blast (http://blast.ncbi.nlm.nih.gov/Blast.cgi) and UCSC Blat browsers (http://genome.ucsc.edu/) to analyze the sequence of those putative trimeric ESTs. Since some of the chimeric, trimeric and tetrameric ESTs contained mt-sequences and since mtDNA was highly polymorphic, different mtDNA sequences (with access number provided if not the UCSC reference) were used as the reference to get the best match during sequence alignment. Therefore, the exact nt numbers and locations might slightly differ from different publications.

RT-PCR, T-A cloning and DNA sequencing

We cultured HEK293 and Hela cells in the routine fashion and extracted total RNA from whole cell lysates with Trizol as described before [36,37]. In some experiments, RNA samples were treated with Ambion® TURBO DNA-free™ Kit (Cat # AM1907, Life Technology), which has a much higher affinity to DNA than conventional DNase I, to remove DNA. RT of the total RNA to cDNA was performed using random hexamers, followed by PCR to amplify cDNA with primers specific to different ESTs. The novel, spliced mtRNA 16SrRNA-s was amplified with mtF2006 (5′-GGTGATAGCTGGTTGTCCAA-3′) as the forward primer and mtR3071 (5′-GAACTCAGATCACGTAGGAC-3′) as the reverse primer. All PCR products were fractioned in 0.9–1.2% agarose gel and visualized by ethidium bromide staining. The band of interest was excised out and purified using a simple method we described before [38]. The purified DNA fragment was cloned into a T-A vector and the resultant plasmids were sent to Genewiz (http://www.genewiz.com) for sequencing. The details of these methods were described before [15,37].

Acknowledgments

We want to thank Fred Bogott, M.D., Ph.D., at Austin Medical Center, Austin, Minnesota, for his excellent English editing of this manuscript.

Author Contributions

Conceived and designed the experiments: WY DJL. Performed the experiments: GWC YCOY. Analyzed the data: ADB EPH EW JHZ JMW DJL HHS. Wrote the paper: WY DJL.

References

1. Celestino R, Sigstad E, Lovf M, Thomassen GO, Groholt KK, et al. (2012) Survey of 548 oncogenic fusion transcripts in thyroid tumors supports the importance of the already established thyroid fusions genes. Genes Chromosomes Cancer 51: 1154–1164.
2. Lasda EL, Blumenthal T (2011) Trans-splicing. Wiley Interdiscip Rev RNA 2: 417–434.
3. Pink JJ, Wu SQ, Wolf DM, Bilimoria MM, Jordan VC (1996) A novel 80 kDa human estrogen receptor containing a duplication of exons 6 and 7. Nucleic Acids Res 24: 962–969.
4. Pink JJ, Fritsch M, Bilimoria MM, Assikis VJ, Jordan VC (1997) Cloning and characterization of a 77-kDa oestrogen receptor isolated from a human breast cancer cell line. Br J Cancer 75: 17–27.
5. Birney E, Stamatoyannopoulos JA, Dutta A, Guigo R, Gingeras TR, et al. (2007) Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature 447: 799–816.
6. Gingeras TR (2009) Implications of chimaeric non-co-linear transcripts. Nature 461: 206–211.
7. Li X, Zhao L, Jiang H, Wang W (2009) Short homologous sequences are strongly associated with the generation of chimeric RNAs in eukaryotes. J Mol Evol 68: 56–65.
8. Yang M, Wu J, Wu SH, Bi AD, Liao DJ (2012) Splicing of mouse p53 pre-mRNA does not always follow the "first come, first served" principle and may be influenced by cisplatin treatment and serum starvation. Mol Biol Rep -DOI: 10.1007/s11033-012-1798-2.
9. Ramos A, Barbena E, Mateiu L, del Mar GM, Mairal Q, et al. (2011) Nuclear insertions of mitochondrial origin: Database updating and usefulness in cancer studies. Mitochondrion 11: 946–953.
10. Tsuji J, Frith MC, Tomii K, Horton P (2012) Mammalian NUMT insertion is non-random. Nucleic Acids Res 40: 9073–9088.
11. Nicholls TJ, Rorbach J, Minczuk M (2013) Mitochondria: Mitochondrial RNA metabolism and human disease. Int J Biochem Cell Biol 45: 845–849.
12. Rorbach J, Minczuk M (2012) The post-transcriptional life of mammalian mitochondrial RNA. Biochem J 444: 357–373.
13. Seligmann H (2012) Overlapping genes coded in the 3′-to-5′-direction in mitochondrial genes and 3′-to-5′ polymerization of non-complementary RNA by an 'invertase'. J Theor Biol 315: 38–52.
14. Seligmann H (2013) Systematic asymmetric nucleotide exchanges produce human mitochondrial RNAs cryptically encoding for overlapping protein coding genes. J Theor Biol 324: 1–20.
15. Yuan C, Liu Y, Yang M, Liao DJ (2013) New methods as alternative or corrective measures for the pitfalls and artifacts of reverse transcription and polymerase chain reactions (RT-PCR) in cloning chimeric or antisense-accompanied RNA. RNA Biol 10: 958–968.
16. Mercer TR, Neph S, Dinger ME, Crawford J, Smith MA, et al. (2011) The human mitochondrial transcriptome. Cell 146: 645–658.
17. Rackham O, Mercer TR, Filipovska A (2012) The human mitochondrial transcriptome and the RNA-binding proteins that regulate its expression. Wiley Interdiscip Rev RNA 3: 675–695.
18. Landerer E, Villegas J, Burzio VA, Oliveira L, Villota C, et al. (2011) Nuclear localization of the mitochondrial ncRNAs in normal and cancer cells. Cell Oncol (Dordr) 34: 297–305.
19. Bandiera S, Ruberg S, Girard M, Cagnard N, Hanein S, et al. (2011) Nuclear outsourcing of RNA interference components to human mitochondria. PLoS One 6: e20746, doi: 10.1371/journal.pone.0020746.
20. Das S, Ferlito M, Kent OA, Fox-Talbot K, Wang R, et al. (2012) Nuclear miRNA regulates the mitochondrial genome in the heart. Circ Res 110: 1596–1603.
21. Charlesworth A, Meijer HA, de Moor CH (2013) Specificity factors in cytoplasmic polyadenylation. Wiley Interdiscip Rev RNA 4: 437–461.
22. Farre JC, Aknin C, Araya A, Castandet B (2012) RNA editing in mitochondrial trans-introns is required for splicing. PLoS One 7: e52644, doi: 10.1371/journal.pone.0052644.
23. Jackson CJ, Waller RF (2013) A widespread and unusual RNA trans-splicing type in dinoflagellate mitochondria. PLoS One 8: e56777, doi: 10.1371/journal.pone.0056777.
24. Grewe F, Herres S, Viehover P, Polsakiewicz M, Weisshaar B, et al. (2011) A unique transcriptome: 1782 positions of RNA editing alter 1406 codon identities in mitochondrial mRNAs of the lycophyte Isoetes engelmannii. Nucleic Acids Res 39: 2890–2902.
25. Szczesny RJ, Wojcik MA, Borowski LS, Szewczyk MJ, Skrok MM, et al. (2013) Yeast and human mitochondrial helicases. Biochim Biophys Acta 1829: 842–53.
26. Burzio VA, Villota C, Villegas J, Landerer E, Boccardo E, et al. (2009) Expression of a family of noncoding mitochondrial RNAs distinguishes normal from cancer cells. Proc Natl Acad Sci U S A 106: 9430–9434.
27. Villegas J, Burzio V, Villota C, Landerer E, Martinez R, et al. (2007) Expression of a novel non-coding mitochondrial RNA in human proliferating cells. Nucleic Acids Res 35: 7336–7347.
28. Villegas J, Zarraga AM, Muller I, Montecinos L, Werner E, et al. (2000) A novel chimeric mitochondrial RNA localized in the nucleus of mouse sperm. DNA Cell Biol 19: 579–588.
29. Wallace DC (2010) Mitochondrial DNA mutations in disease and aging. Environ Mol Mutagen 51: 440–450.
30. Aral C, Akkiprik M, Kaya H, taizi-Celikel C, Caglayan S, et al. (2010) Mitochondrial DNA common deletion is not associated with thyroid, breast and colorectal tumors in Turkish patients. Genet Mol Biol 33: 1–4.
31. Tonska K, Piekutowska-Abramczuk D, Kaliszewska M, Kowalski P, Tanska A, et al. (2012) Molecular investigations of mitochondrial deletions: evaluating the usefulness of different genetic tests. Gene 506: 161–165.
32. Woischnik M, Moraes CT (2002) Pattern of organization of human mitochondrial pseudogenes in the nuclear genome. Genome Res 12: 885–893.
33. Veltri KL, Espiritu M, Singh G (1990) Distinct genomic copy number in mitochondria of different mammalian organs. J Cell Physiol 143: 160–164.
34. Lebedeva MA, Shadel GS (2007) Cell cycle- and ribonucleotide reductase-driven changes in mtDNA copy number influence mtDNA Inheritance without compromising mitochondrial gene expression. Cell Cycle 6: 2048–2057.
35. Alexeyev M, Shokolenko I, Wilson G, Ledoux S (2013) The maintenance of mitochondrial DNA integrity – critical analysis and update. Cold Spring Harb Perspect Biol 5:a012641, doi: 10.1101/cshperspect.a012641.
36. Sun Y, Li Y, Luo D, Liao DJ (2012) Pseudogenes as Weaknesses of ACTB (Actb) and GAPDH (Gapdh) Used as Reference Genes in Reverse Transcription and Polymerase Chain Reactions. PLoS One 7: e41659–doi:10.1371/journal.-pone.0041659.
37. Yang M, Sun Y, Ma L, Wang C, Wu JM, et al. (2011) Complex alternative splicing of the smarca2 gene suggests the importance of smarca2-B variants. J Cancer 2: 386–400.
38. Sun Y, Sriramajayam K, Luo D, Liao DJ (2012) A quick, cost-free method of purification of DNA fragments from agarose gel. J Cancer 3: 93–95.
39. Sun Y, Cao S, Yang M, Wu S, Wang Z, et al. (2013) Basic anatomy and tumor biology of the RPS6KA6 gene that encodes the p90 ribosomal S6 kinase-4. Oncogene 32: 1794–1810.
40. Yang M, Wu J, Wu SH, Bi AD, Liao DJ (2012) Splicing of mouse p53 pre-mRNA does not always follow the "first come, first served" principle and may be influenced by cisplatin treatment and serum starvation. Mol Biol Rep 39: 9247–9256.